Data Warehousing & ITS

Using Archived Data to Improve Transportation Services

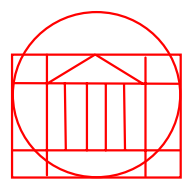
Smart Travel Laboratory

Catherine C. McGhee
Virginia Transportation Research Council

Brian L. Smith
University of Virginia

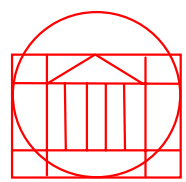
Presentation Overview

- Virginia Smart Travel Laboratory
- ITS Data Warehousing
  - Database size/structure
  - Data screening
  - Data extraction
- Extracting *information* from data
  - Prototype tools
  - Analysis



Smart Travel Laboratory

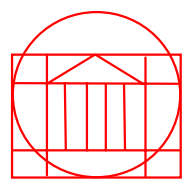
- Laboratory established specifically to support the development and operations of ITS
- Joint Facility: University of Virginia and VA Transportation Research Council
- Distinguishing characteristics
  - Access to real-time ITS data and video
  - Advanced information technology




Bi-directional traffic

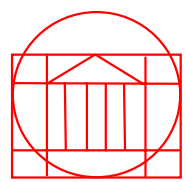
Research Program Focus

- Application of advanced information technology to surface transportation.
- Derive *information* from *data* to support *intelligent* transportation decision making.
- Interdisciplinary approach



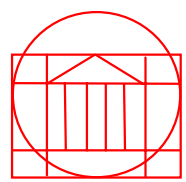
Phase I - Complete

- Integration with Hampton Roads Smart Traffic Center
  - Freeway management system
  - 203 sensors stations
  - 38 video cameras
- Integration with Northern Virginia Traffic Signal System
  - Over 900 intersections



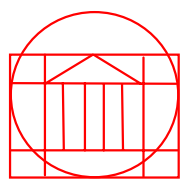
Phase II - Winter 2000

- Integration with Northern Virginia Smart Traffic Center
  - 100 CCTV cameras
  - 200 variable message signs
- Integration with Richmond Smart Traffic Center
- Integration with US Wireless Corp's network operations center



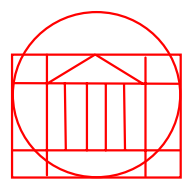
Bottom Line

We have an awful lot of data - what should be done with all of it??



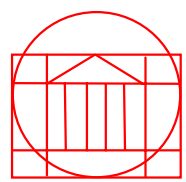
STL Data Management

- Oracle 8 DBMS
- 40 gigabytes and growing by the minute!
- Over 30 tables to support multiple applications
  - Reflects creation of numerous data marts to support real-time application requirements.



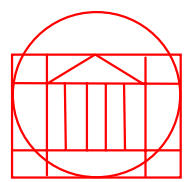
Data Quality

- Just because it is ITS data doesn't mean that it is good data.
- In general, 25% of the data we "should" receive is missing or erroneous.
- We have developed a number of data screening techniques.
- Current research -- estimate data based on surrounding sensors to fill in the gaps.



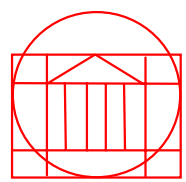
Data Screening

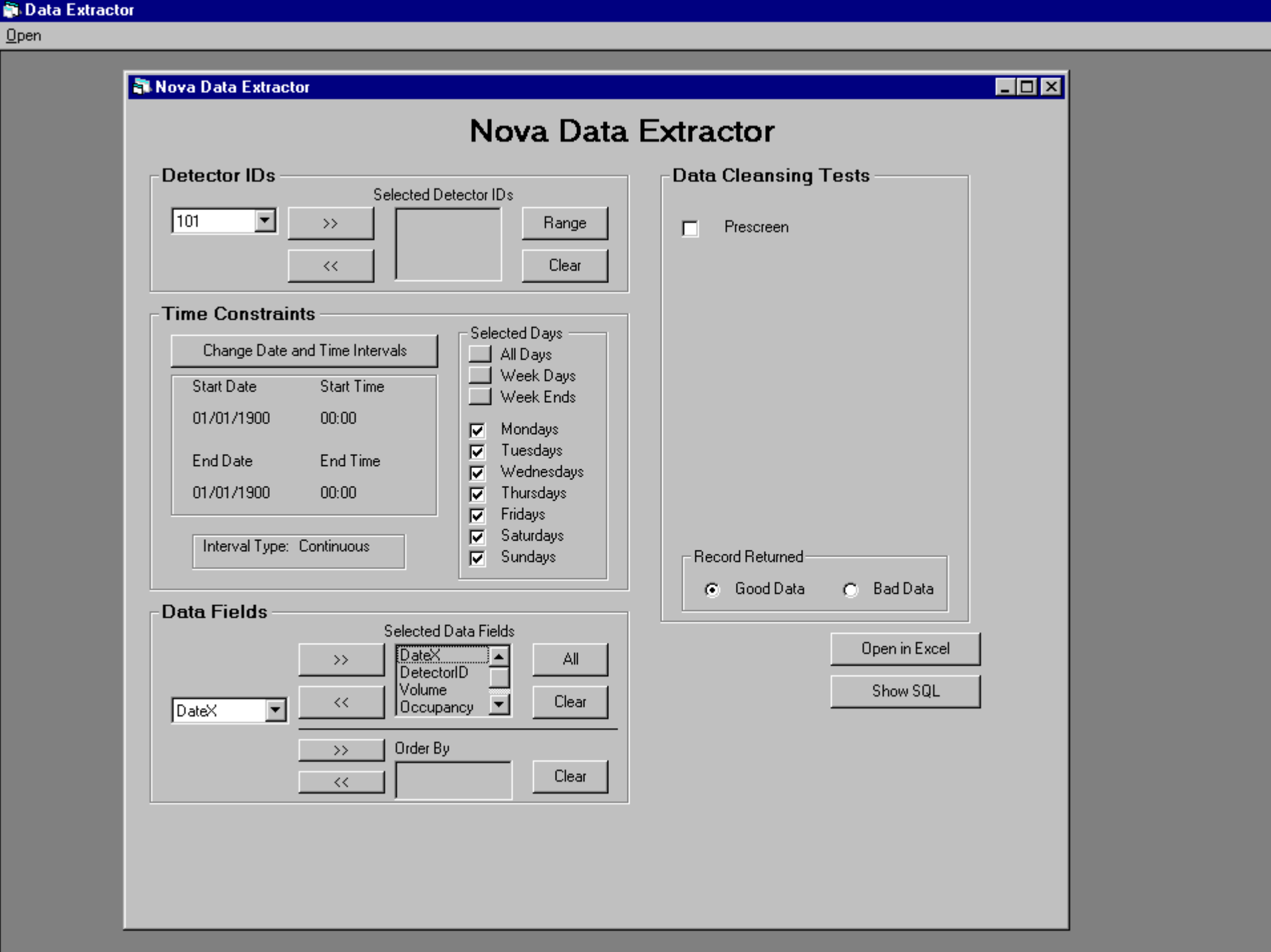
- Most typical data screening approach - thresholds
  - Volume, speed, or occupancy greater or less than a threshold value
- We have had success using an average effective vehicle length test - simultaneously using volume, speed, and occupancy.
  - TRB 2000 paper (Turochy & Smith)

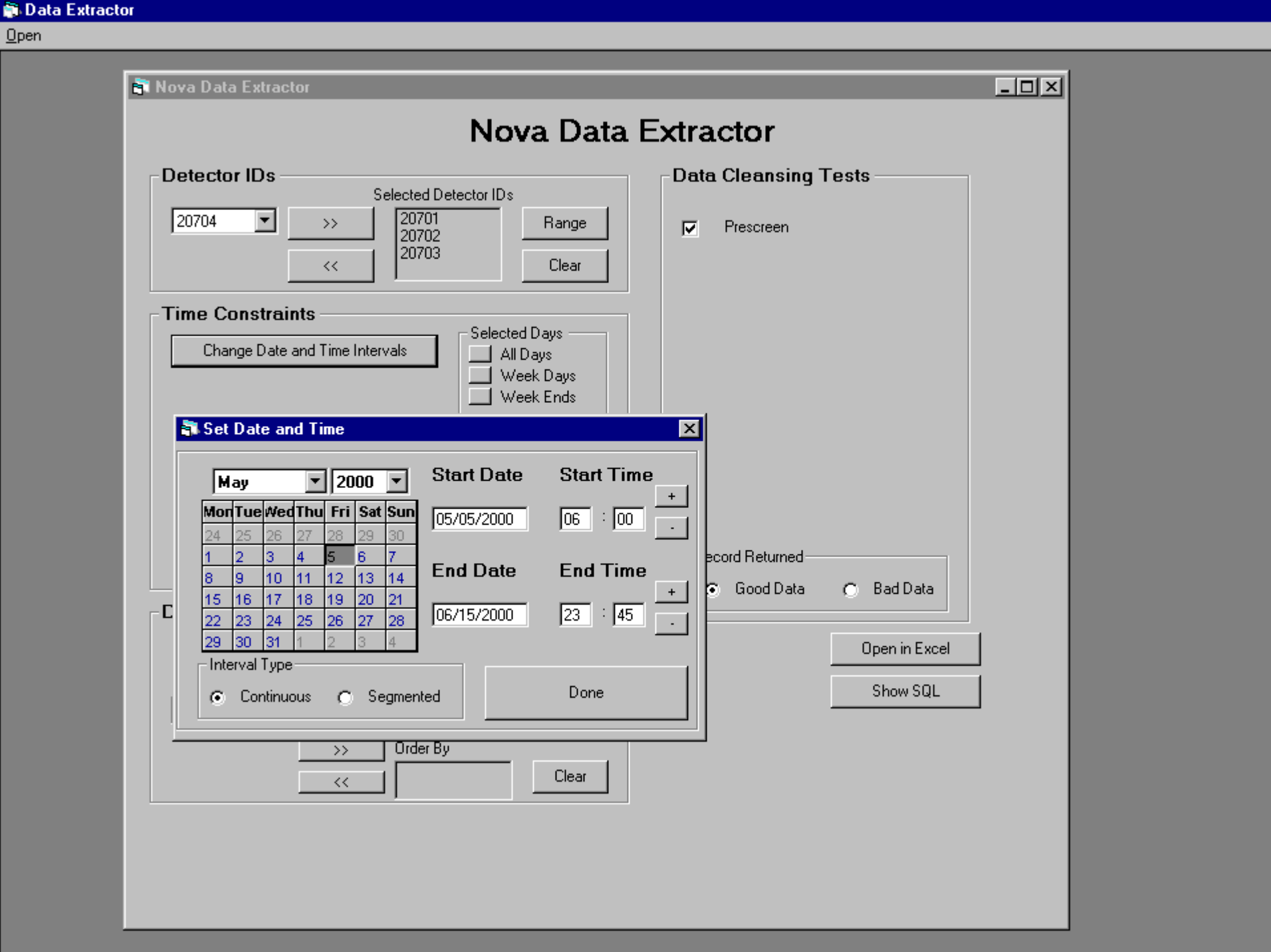


Data Extraction

- Easy access to tailored data sets required for analysis
  - Location, date/time selection
  - Compilation interval selectable
  - Screening
- Extraction tools developed
  - Used by Virginia and Wisconsin DOT's
- Currently working to web-enable

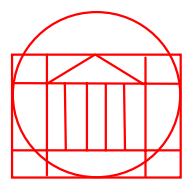



Data Extractor
Open
Nova Data Extractor
Nova Data Extractor
Detector IDs
Selected Detector IDs
101
>>
<<
Range
Clear
Time Constraints
Change Date and Time Intervals
Start Date
Start Time
01/01/1900
00:00
End Date
End Time
01/01/1900
00:00
Interval Type: Continuous
Selected Days
All Days
Week Days
Week Ends
Mondays
Tuesdays
Wednesdays
Thursdays
Fridays
Saturdays
Sundays
Data Fields
Selected Data Fields
DateX
>>
<<
DateX
DetectorID
Volume
Occupancy
All
Clear
>>
<<
Order By
Clear
Data Cleansing Tests
Prescreen
Record Returned
Good Data
Bad Data
Open in Excel
Show SQL


Data Extractor
Open
Nova Data Extractor
Nova Data Extractor
Detector IDs
20704
>>
<<
Selected Detector IDs
20701
20702
20703
Range
Clear
Data Cleansing Tests
Prescreen
Time Constraints
Change Date and Time Intervals
Selected Days
All Days
Week Days
Week Ends
Set Date and Time
May
2000
Mon Tue Wed Thu Fri Sat Sun
24 25 26 27 28 29 30
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 1 2 3 4
Start Date
05/05/2000
Start Time
06 : 00
End Date
06/15/2000
End Time
23 : 45
+
-
Interval Type
Continuous
Segmented
Done
ecord Returned
Good Data
Bad Data
Open in Excel
Show SQL
>>
<<
Order By
Clear

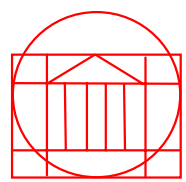
Extracting *Information* from Data

- This is why we collect all of this data in the first place!
- Prototype tools
  - Traffic flow forecasting
  - Signal timing interval selection support
- Analysis
  - Speed-flow relationships
  - Flow rate measurement interval

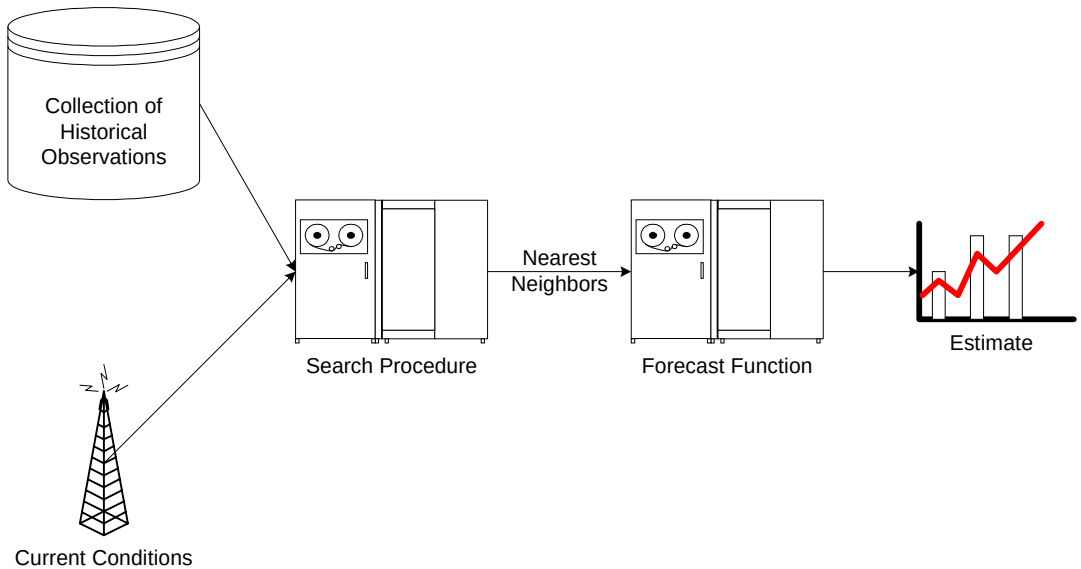


Traffic Flow Forecasting

- ITS must be able to predict future traffic conditions in a timely manner and take appropriate actions based on these forecasts
- Traffic flow forecasting research program
  - On-line real-time forecasts that take advantage of archived data
  - Nonparametric regression - preferred approach



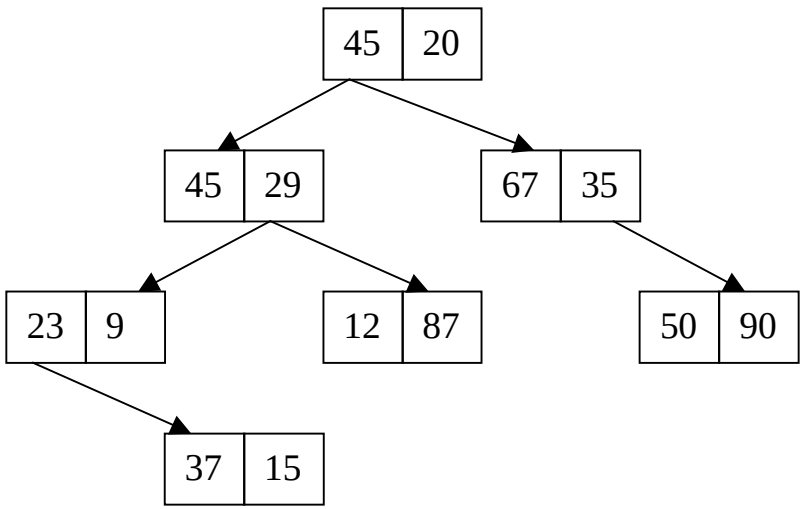
Nonparametric Regression (NPR)



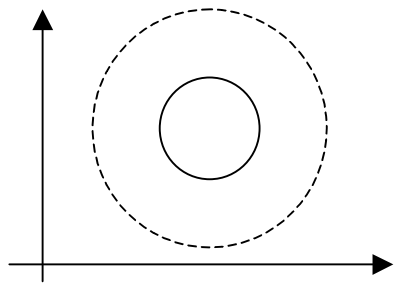
- Forecasting technique similar to case-based reasoning
  - Searches a collection of historical observations for past cases similar to the current conditions to generate a forecast
- Executes slowly

Methods to Speed NPR

- Advanced data structures
  - Multidimensional binary search trees

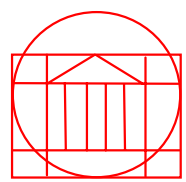


- Approximate nearest neighbors
  - Use historical data points sufficiently close to the query point but are not necessarily the closest points



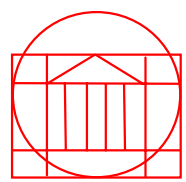
Research Results

- Use of advanced data structures alone reduces execution time by a factor of 1,000
- Forecasting prototype now meets real-time requirements and produces results with roughly 10% error.
- On-line forecasts available on Smart Travel Laboratory website.

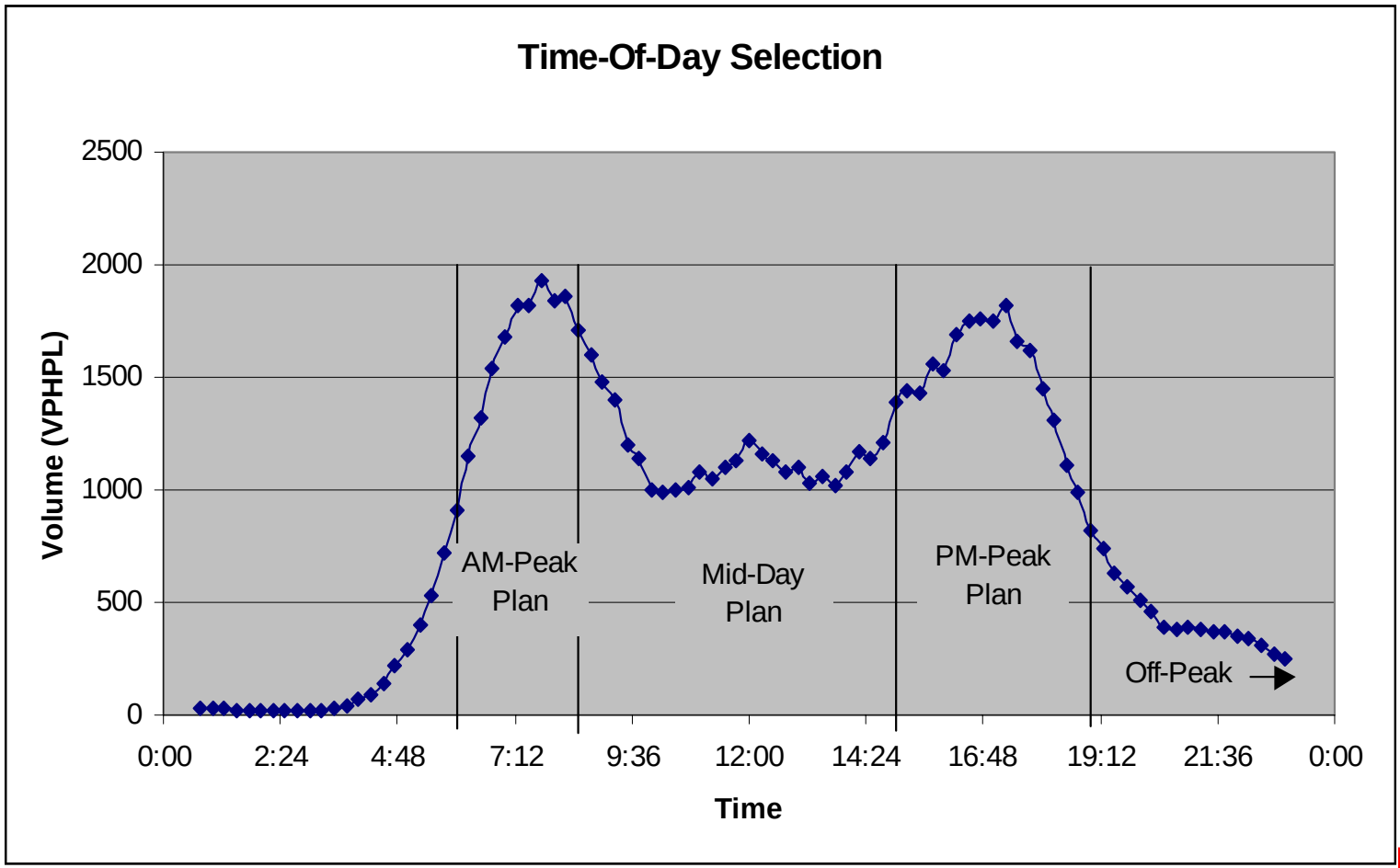


Timing Plan Development Assistance Tool

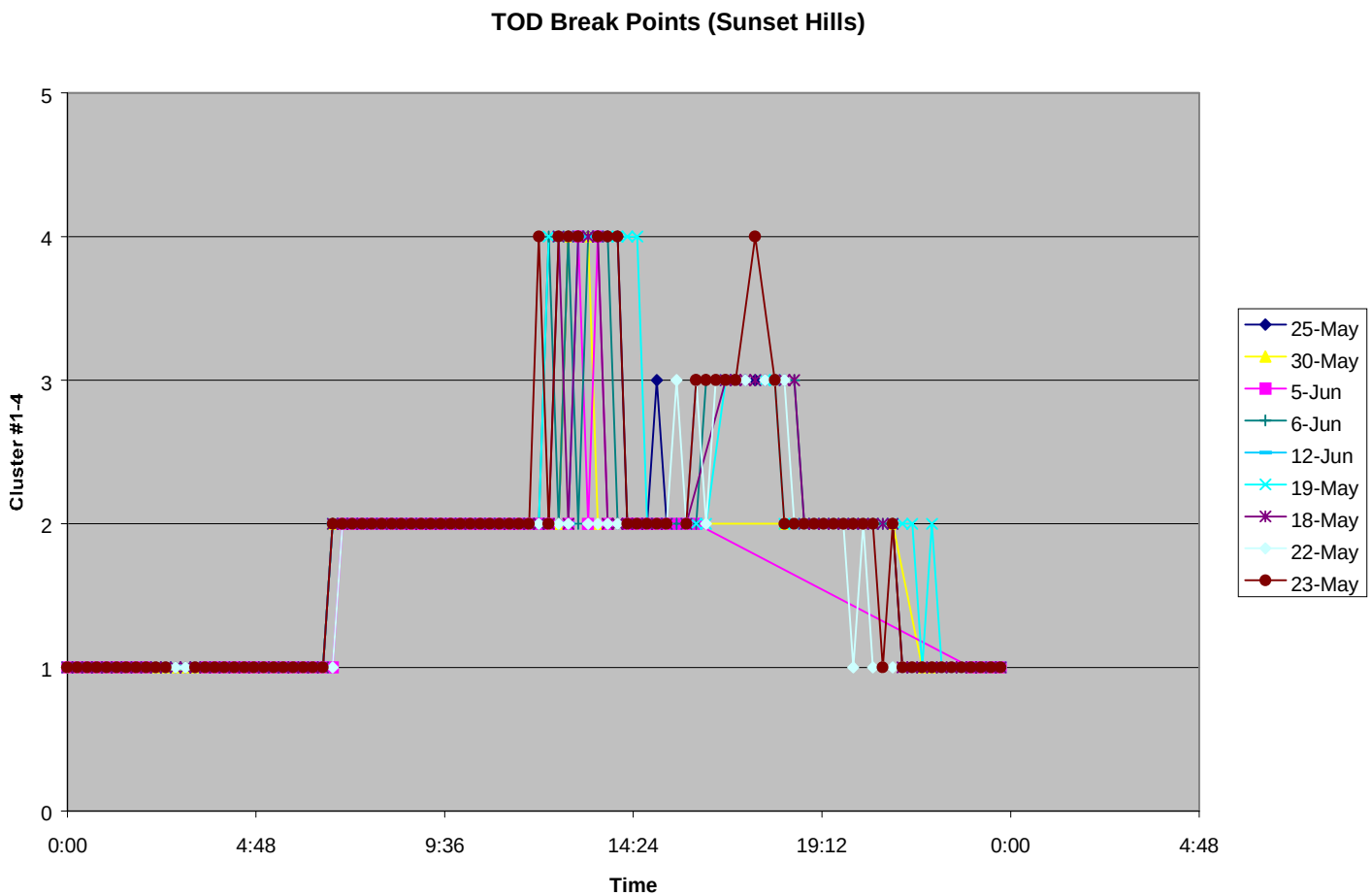
- Explore “state” definition for a corridor
- Use statistical clustering techniques to identify ideal “break-points” between time-of-day periods.
- If successful, this would serve as the basis for a tool that would supplement existing tools (such as SYNCHRO) in timing plan development.



Current Practice



New Break Points



Comparison

AM-Peak Plan (06:30 – 10:00)

Mid-Day Plan (10:00 – 15:00)

PM-Peak Plan (15:00 – 19:30)

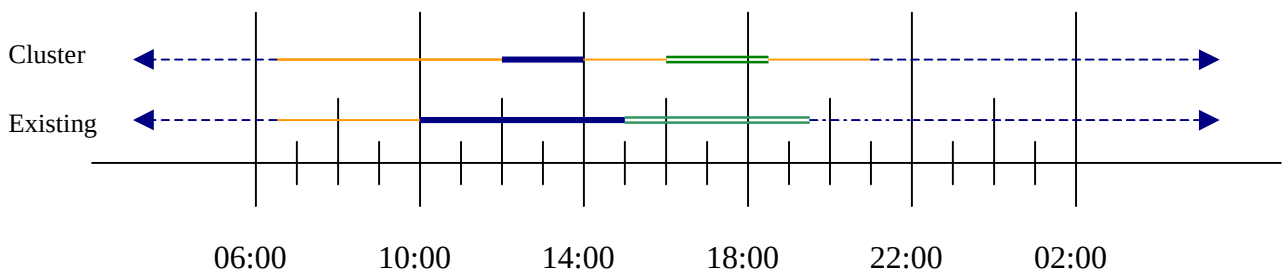
Off-Peak Plan (19:30 – 06:30)

AM-Peak Plan (06:45 – 11:45, 14:00 – 15:45, 18:30 – 21:00)

Mid-Day Plan (11:45 – 14:00)

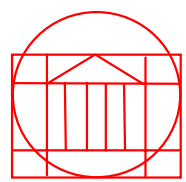
PM-Peak Plan (15:45 – 18:30)

Off-Peak Plan (21:00 – 06:45)

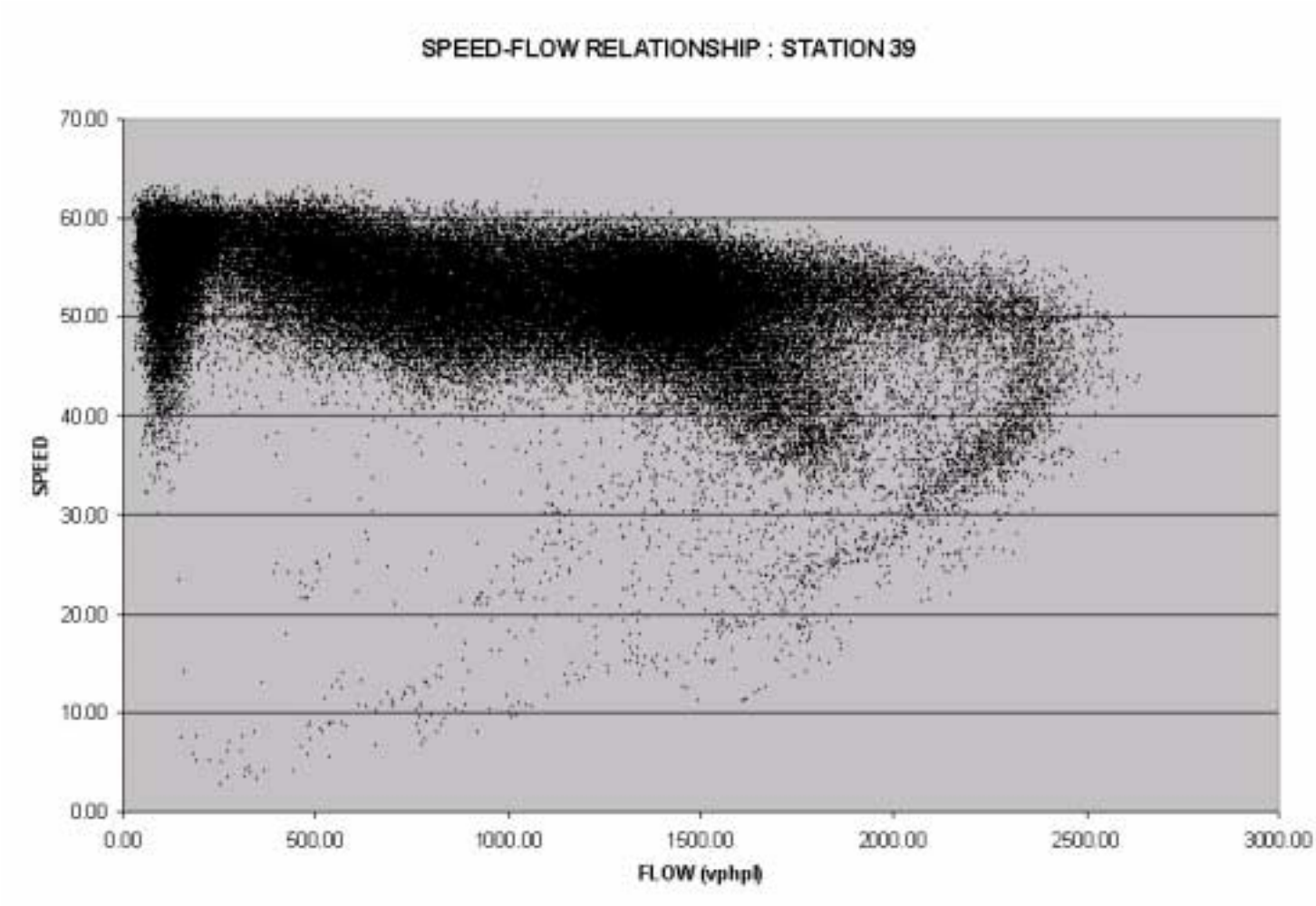


Speed-Flow Relationships

- It is important to understand how particular segments of the freeway network operate
- *Highway Capacity Manual* provides general guidance based on limited data
- Use ITS data to address this issue on a site-specific basis.

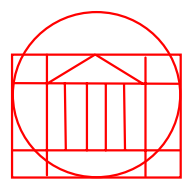


Example


SPEED-FLOW RELATIONSHIP : STATION 39
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
SPEED
0.00
500.00
1000.00
1500.00
2000.00
2500.00
3000.00
FLOW (vphpl)

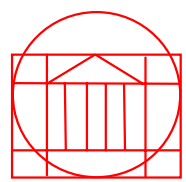
Flow Rate Measurement Interval

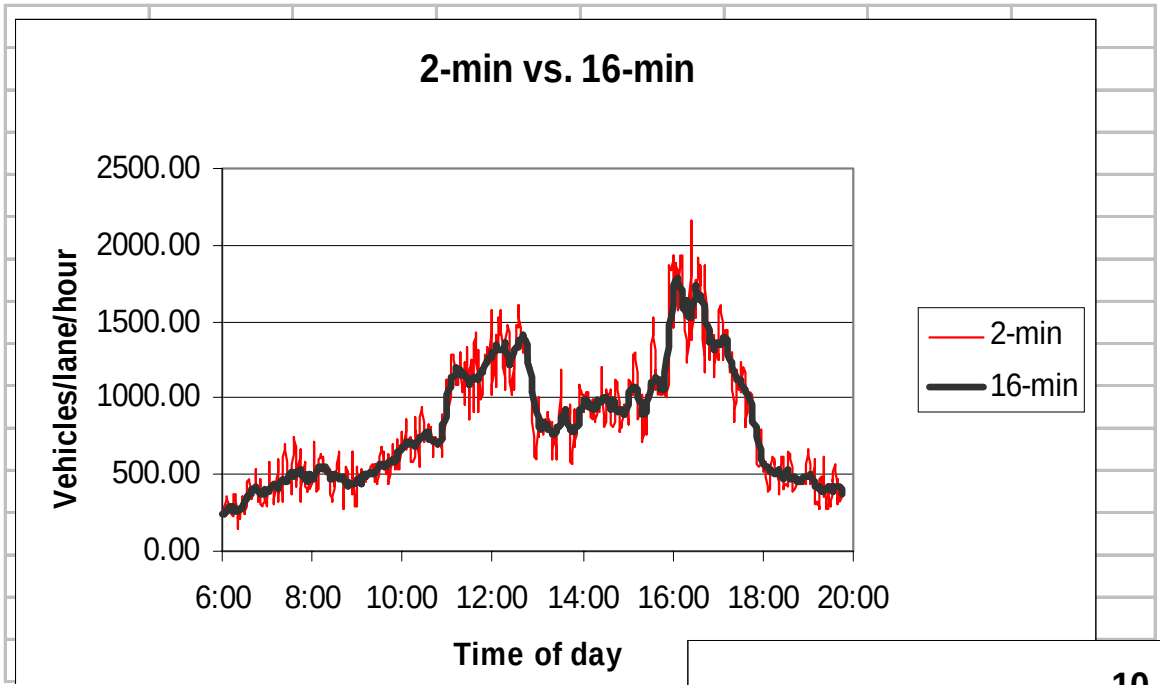
- Highway Capacity Manual states: “5-minute flow rates have been avoided, since research has shown them to be statistically unstable”
- Suggested interval - 15 minutes
- Problematic for ITS applications (and others) -- a lot can change in 15 minutes

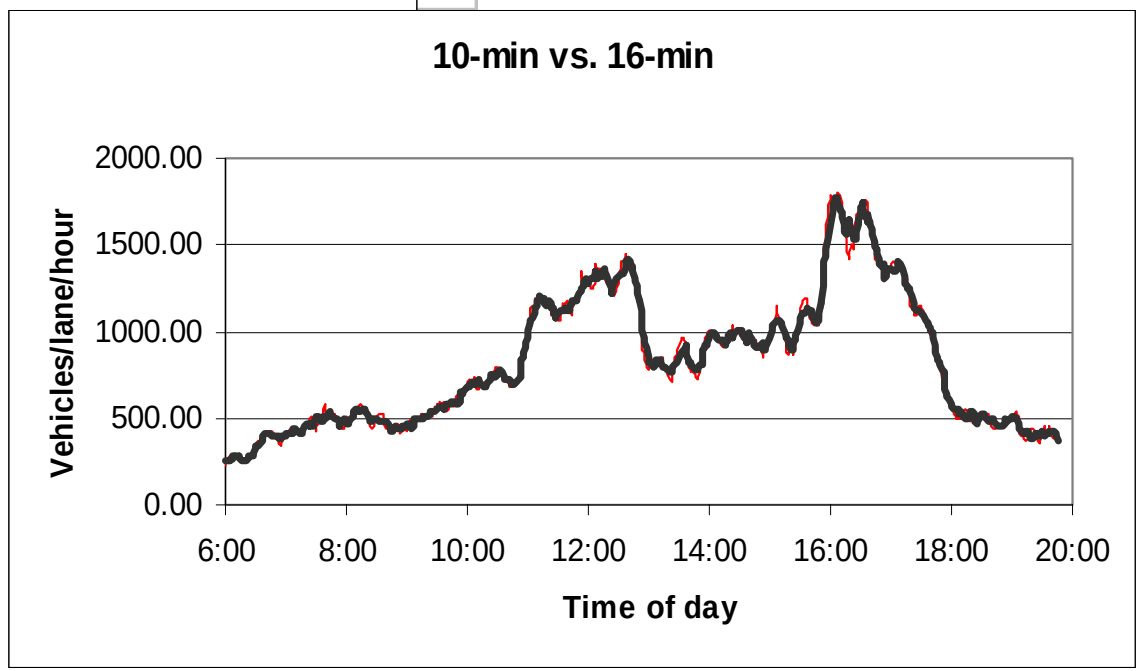


Investigate Concept

- Consider flow rates measured at different time intervals (2, 4, 6, 8, 10, 12, 14, 16 minutes) -- how does interval impact “noise” in the signal?
- Allows transportation professionals to better understand the data they are working with.

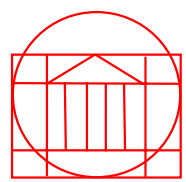



2-min vs. 16-min
2500.00
2000.00
1500.00
1000.00
500.00
0.00
Vehicles/lane/hour
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
Time of day
2-min
16-min


10-min vs. 16-min
2000.00
1500.00
1000.00
500.00
0.00
Vehicles/lane/hour
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
Time of day

Conclusions

- ITS provides the transportation profession with a tremendous new source of data
- In order to make use of this data, research and development is needed:
  - data management
  - decision support tools
  - analysis techniques



For More Information

http://SmartTravelLab.virginia.edu

Cathy C. McGhee
McGheeCC@vdot.state.va.us

Brian L. Smith
briansmith@virginia.edu